

## 1. Pendahuluan

### Latar Belakang

*Video game* telah menjadi salah satu bentuk hiburan paling populer di dunia, didorong oleh pesatnya pertumbuhan industri teknologi digital. Perusahaan pengembang teknologi digital berlomba-lomba untuk meraih keuntungan besar dari industri ini. Oleh karena itu, sangat penting bagi perusahaan *game* untuk memahami kebutuhan para penggemar *video game*. Pendapat konsumen dalam ulasan *game* memberikan informasi yang berharga bagi perusahaan, karena ulasan tersebut menjadi dasar penting dalam pengembangan produk *game*. Selain itu, ulasan terkait *game* juga mempengaruhi calon konsumen, karena ulasan tersebut menjadi bahan pertimbangan utama untuk menjawab kekhawatiran konsumen tentang kualitas *game* [1].

Salah satu pemain utama dalam industri *video game* saat ini adalah Steam, sebuah layanan distribusi *video game* yang dikembangkan oleh *Valve Corporation* dan telah beroperasi sejak tahun 2003. Pengaruh *Steam* dalam distribusi *game* mencakup lebih dari 30.000 *game* yang dinikmati oleh 18 juta pengguna aktif pada platform digital *Steam*. Berdasarkan statistik layanan digital, *Steam* menguasai lebih dari separuh pangsa pasar penjualan *game* komputer secara global. Dengan fitur ulasan produk yang disediakan oleh *Steam*, rekomendasi pengguna atau sebaliknya memegang peranan penting dalam membimbing keputusan pengguna lain terkait suatu *game* [2]. Dengan banyaknya data ulasan yang tersedia di *Steam*, penelitian ini berfokus pada skenario analisis sentimen terhadap ulasan pengguna, menggunakan metode klasifikasi *Random Forest* dan ekstraksi fitur *Chi-Square*.

Sentimen dalam ulasan *game* di *Steam* merupakan data yang sangat berharga bagi perusahaan pengembang *game*, karena ulasan yang mengungkapkan sentimen mencerminkan pendapat pengguna yang dapat mempengaruhi calon pembeli lainnya [3]. Nilai sentimen dalam ulasan biasanya dikategorikan sebagai positif, negatif, atau netral. Hal ini memungkinkan dilakukannya analisis sentimen pada ulasan tersebut menggunakan metode klasifikasi.

Terdapat beberapa pendekatan klasifikasi yang dapat digunakan untuk menganalisis sentimen dalam ulasan *game* di *Steam*, salah satunya adalah algoritma klasifikasi *Random Forest*. Algoritma *Random Forest* bekerja dengan memilih secara acak nilai vektor sampel untuk membentuk beberapa *decision tree*, di mana *tree* memiliki nilai sampel yang didistribusikan secara independen dan identik di seluruh *tree* [4]. Dalam kinerja klasifikasi *Random Forest*, akurasi terkadang menjadi masalah. Oleh karena itu, diperlukan optimasi untuk meningkatkan kinerja *Random Forest*. Salah satu cara untuk mencapai akurasi yang lebih tinggi adalah dengan melakukan tuning untuk menemukan parameter terbaik.

Peningkatan model *Random Forest* dengan *GridSearchCV* melibatkan proses *fine-tuning* yang dapat memakan waktu sangat lama karena banyaknya kombinasi yang harus diuji serta sifat pencarian yang menyeluruh untuk semua kombinasi parameter dalam grid [5]. *RandomizedSearchCV* adalah metode yang menggunakan optimasi pencarian acak terhadap parameter yang tersedia. Metode ini lebih efektif dalam mengeksplorasi kombinasi parameter secara acak, menghasilkan hasil yang optimal. *RandomizedSearchCV* lebih efisien dalam hal waktu pemrosesan dibandingkan dengan *GridSearchCV*, terutama saat menangani ruang pencarian parameter yang besar, karena hanya mengeksplorasi sebagian dari semua kemungkinan kombinasi [6].

Metode klasifikasi dengan jumlah fitur dan atribut yang besar dapat mengurangi akurasi, sehingga mempengaruhi efisiensi komputasi. Salah satu solusi untuk mengatasi penurunan akurasi ini adalah dengan menggunakan seleksi fitur [6]. Seleksi fitur *Chi-Square* dapat meningkatkan akurasi klasifikasi dengan mengurangi jumlah atribut yang tidak relevan [7]. Selain itu, data ulasan akan diperoleh melalui *web scraping* dari ulasan *Steam* untuk *game Disco Elysium*, dengan tujuan melakukan klasifikasi analisis sentimen menggunakan algoritma *Random Forest* dengan seleksi fitur *Chi-Square*.

### Topik dan Batasannya

Berdasarkan latar belakang yang sudah dijabarkan sebelumnya, didapatkan rumusan masalah sebagai berikut :

1. Bagaimana cara menganalisis sentimen ulasan *game Disco Elysium* menggunakan metode *Random Forest tuning* dengan fitur seleksi *Chi-Square*?
2. Bagaimana hasil akurasi performansi dari analisis sentimen dengan menggunakan metode *Random Forest tuning* dengan fitur seleksi *Chi-Square*?

Dalam penelitian ini dibutuhkan batasan masalah agar penelitian dapat dilaksanakan sesuai dengan yang diharapkan. Adapun batasan masalah sebagai berikut :

1. Ulasan yang digunakan dalam penelitian ini dengan bahasa Inggris yang berasal dari platform penyedia *video game Steam*.
2. Ulasan yang dipilih adalah ulasan yang memiliki kata lebih dari 4 suku kata.
3. Ulasan pengguna yang dipilih dan diproses akan berjumlah 5000 ulasan dengan pelabelan kalimat positif dan kalimat negatif.